# Autonomous Defect Classification in Manufacturing: A Novel Few-Shot Vision-Language Modeling Approach

**Hoang Long Dinh**
Faculty of Control and Automation, Electric Power University, Hanoi, Vietnam
hoanglongnra@gmail.com

**Diem Vuong Doan**
Faculty of Control and Automation, Electric Power University, Hanoi, Vietnam
vuongdd@epu.edu.vn

**Ngoc Khoat Nguyen**
Faculty of Control and Automation, Electric Power University, Hanoi, Vietnam
khoatnn@epu.edu.vn (corresponding author)

**Duy Trung Nguyen**
Faculty of Control and Automation, Electric Power University, Hanoi, Vietnam
trungnd@epu.edu.vn

**Huy Hoang Hoang**
Faculty of Control and Automation, Electric Power University, Hanoi, Vietnam
hoanghh0609@gmail.com

*Received: 21 March 2026 | Revised: 4 May 2026 | Accepted: 11 May 2026*

 *DOI: https://doi.org/10.48084/etasr.18859*

## ABSTRACT

**While established Automated Optical Inspection (AOI) systems and contemporary unsupervised deep learning methodologies exhibit high efficacy in anomaly detection, their inherent "black box" nature precludes them from achieving the semantic interpretability required to ascertain the intrinsic classification of a detected defect. This critical deficiency constitutes a significant operational bottleneck, necessitating costly and intensive manual intervention for labeling and expert-driven retraining to accommodate novel defect typologies. This research introduces VLM-AOI, a paradigm shift achieved through the robust integration of large-scale Vision–Language Models (VLMs) into a two-stage hybrid framework. The system strategically initiates with a localization stage, employing a state-of-the-art unsupervised model to precisely delineate candidate defect regions. Subsequently, the interpretation stage leverages the VLM's powerful cross-modal understanding to execute zero-shot classification of these localized regions based on rich semantic text prompts. Experimental outcomes validate that the VLM-AOI framework not only preserves the high pixel-level detection fidelity of current methods but also delivers superior zero-shot classification accuracy, effectively identifying defect classes entirely unrepresented in the training data. This innovation dramatically enhances operational adaptability and reduces both manufacturing downtime and associated expenditure.**

***Keywords-Automated Optical Inspection (AOI); Vision–Language Model (VLM); zero-shot defect classification; smart manufacturing; semantic understanding; anomaly detection***

## I. INTRODUCTION

In the era of Industry 4.0, Automated Optical Inspection (AOI) plays a critical role in quality assurance for high-precision manufacturing sectors such as Printed Circuit Boards (PCBs), semiconductors, and pharmaceuticals. Recent advances in deep learning have significantly propelled the capabilities of AOI systems. For instance, state-of-the-art You Only Look Once (YOLO) architectures, such as YOLOv10,

have been optimized for high-speed, multi-class PCB defect detection by refining feature extraction layers [1].

Similarly, advanced Artificial Neural Networks (ANNs) have been successfully deployed to isolate complex textural anomalies in fabric manufacturing, proving the versatility of deep learning across different industrial substrates [2]. These methodologies, especially unsupervised anomaly detection paradigms, have significantly improved AOI by enabling accurate defect detection even under limited data conditions.

State-of-the-art methods such as PatchCore [3] and PaDiM [4] model the normal state of a system and detect deviations effectively. However, these methods are limited to anomaly detection and localization; they cannot provide semantic understanding of the defect type, such as distinguishing between a scratch and a misplaced component. As a result, manual inspection and classification are still required, and newly discovered defects demand costly data collection and retraining of supervised models [5].

Traditional supervised Convolutional Neural Network (CNN)-based classifiers, such as ResNet and VGG, also lack flexibility because they require large, labeled datasets for each defect class and fail when encountering unseen defect types [6, 7]. A broader overview of zero-shot learning methods is presented in [8].

Meanwhile, Vision–Language Models (VLMs), such as Contrastive Language–Image Pretraining (CLIP) [9] and A Large-scale ImaGe and Noisy-text embedding (ALIGN) [10], have demonstrated strong zero-shot classification ability by learning joint visual-textual representations from large-scale image-text pairs. However, their application in industrial environments remains challenging due to the domain gap between internet-scale data and specialized industrial imagery [11, 12], as well as their high computational cost [13].

To address this gap, this paper proposes the Hybrid AOI Framework (VLM-AOI), which combines the speed and localization capability of unsupervised anomaly detection with the semantic generalization ability of VLMs. Unsupervised approaches remain essential in AOI because they require only normal samples [14], whereas recent studies have started exploring alternatives beyond traditional CNNs for PCB defect detection [15]. Methods such as WinCLIP [16], supported by contrastive learning principles [17], demonstrate the potential of zero-/few-shot anomaly classification, but challenges related to latency and deployment in smart factory environments remain.

## II. PROPOSED METHOD (THE VLM-AOI FRAMEWORK)

The research proposes a two-stage architecture, illustrated in Figure 1, which is designed to optimize between processing speed (Stage 1) and in-depth analytical capability (Stage 2). The architectural choice of using a transformer-based backbone for feature extraction is inspired by their recent success in vision tasks.

The provided block diagram in Figure 2 illustrates a robust two-stage AOI-VLM hybrid framework designed for efficient defect analysis. Phase 1 operates as an unsupervised anomaly detection module, rapidly extracting features and generating binary masks to identify and crop potential Regions of Interest (RoIs) from the test image.

If RoIs are detected, the system proceeds to Phase 2, where the cropped anomaly patches undergo VLM-based processing, involving image and text encoding, to perform accurate zero-shot defect classification. This architecture optimizes inspection by dedicating high-cost VLM analysis only to the relevant, fault-containing regions identified in the initial phase.

### *A. High-Speed Anomaly Region Localization*

This stage functions as an intelligent and fast "filter," relieving the VLM from the burden of analyzing the entire high-resolution image (often 2K–4K resolution). Instead, it first identifies potential anomalous RoIs. PatchCore was selected as the foundation for this stage due to its state-of-the-art performance and high inference speed.

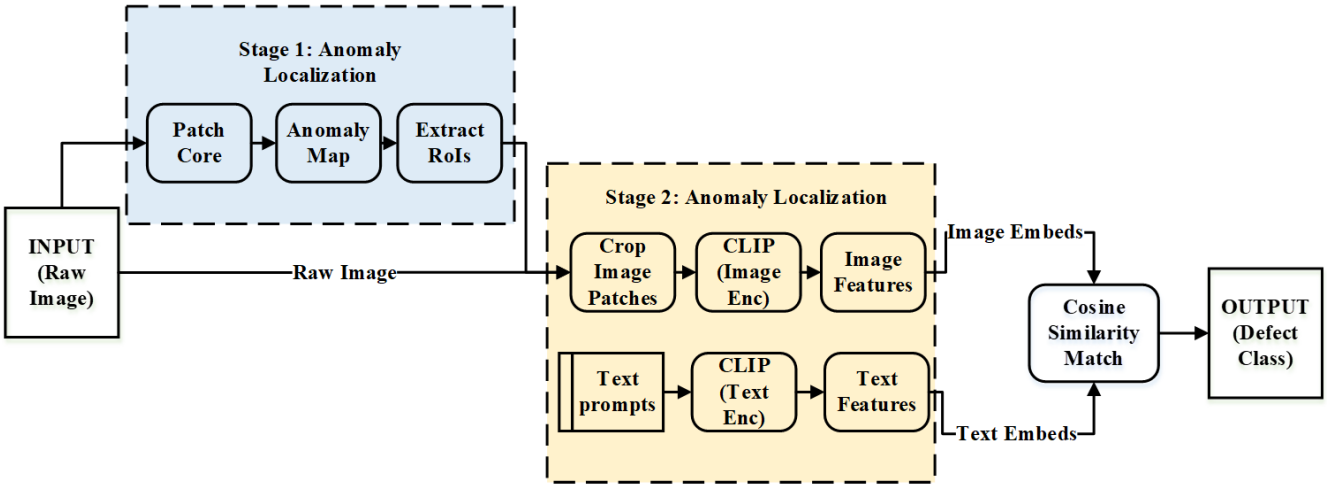


Fig. 1. Proposed two-stage VLM-AOI architecture.

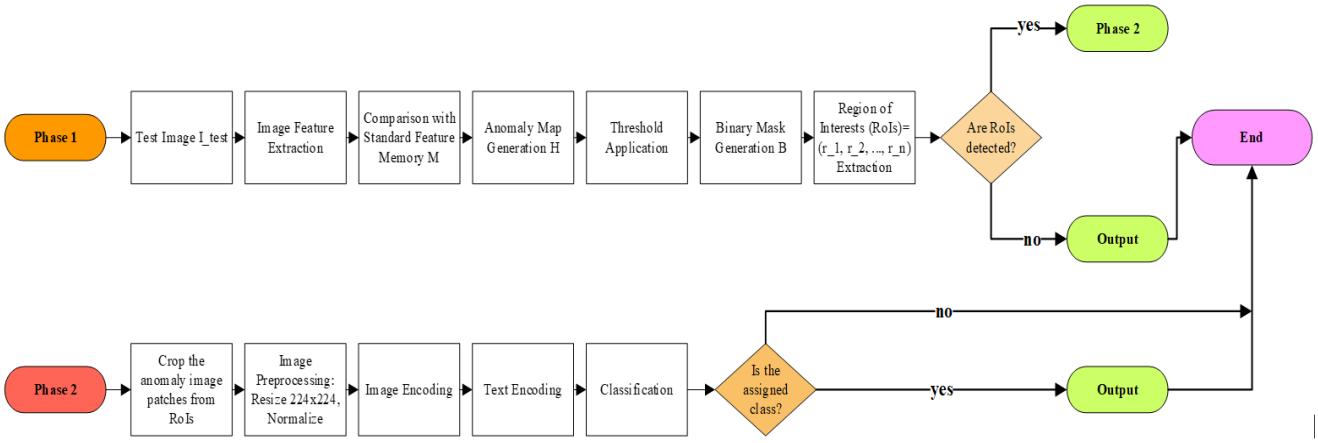


Fig. 2. System block diagram of the AOI-VLM architecture.

During training, the model is exposed only to a set of "normal" images, $I_{train}$, from which PatchCore constructs a "memory bank," $M$, of features representing the normal state. During the inference phase, a test image $I_{test}$ is processed by extracting its patch features and comparing them against the memory bank $M$. This comparison yields an anomaly map, $H$, in which each pixel value represents its anomaly score. A threshold, $\theta$ (typically determined using a validation set), is then applied to $H$ to generate a binary mask $M_{bin}$. Finally, the connected components extracted from $M_{bin}$ are used to produce a set of bounding boxes for the candidate anomalous regions, $RoIs = \{r_1, r_2, ..., r_n\}$.

### *B. Zero-Shot Semantic Interpretation*

This stage represents the core innovation of the proposed framework. Given the set of RoIs obtained from the previous stage, a VLM is leveraged to semantically interpret the detected anomalies. First, the image regions $r_i$ corresponding to the RoIs are cropped from the original image $I_{test}$ and subjected to basic preprocessing operations, such as resizing to 224 × 224 and normalization, to satisfy the input requirements of CLIP.

Next, "Industrial Prompt Engineering" is employed by defining a "defect library" $L$ in natural language, which includes all defect categories relevant to Quality Assurance (QA) engineers. Critically, this also includes one or more "negative prompts" (e.g., "an image of a normal area" or "an image of a normal circuit board") to effectively suppress false positives generated during Stage 1. For example, a prompt set for PCB inspection might include $T_1$: "an image of a crack on the board," $T_2$: "an image of a missing component," and $T_3$: an image of excess solder."

The CLIP model (e.g., ViT-B/32), which comprises an Image Encoder ($E_I$) and a Text Encoder ($E_T$) is then used to encode both visual and textual information. Each defect patch, $r_i$, is passed through $E_I$ to obtain its image feature vector $v_i$, whereas the entire defect library $L$, is passed through $E_T$ to generate the corresponding set of text feature vectors, $V_T = \{v_{t,1}, v_{t,2}, ..., v_{t,k}\}$. Zero-shot classification is performed by computing the cosine similarity between the image vector $v_i$ and all vectors in $V_T$. The defect class is assigned based on the highest similarity score, as shown in (1):

$$Class(r_i) = \arg\max x_{j\in[1,k]}\left(\frac{v_i . v_{t,j}}{\|v_i\| . \|v_{t,j}\|}\right) \quad (1)$$

If the predicted class, $Class(r_i)$, corresponds to one of the negative prompts, the initial anomaly alert generated during Stage 1 is filtered out, significantly reducing the overall false positive rate.

## III. EXPERIMENTATION AND EVALUATION

This section details the experimental setup and presents a comprehensive analysis of the quantitative and qualitative results of the VLM-AOI framework. The primary focus is the evaluation of Stage 2 (Defect Classification), which constitutes the main contribution of this research.

### *A. Experimental Setup*

For experimentation, the DeepPCB dataset [18], a large-scale publicly available PCB benchmark introduced in [19], was employed. The dataset provides six distinct semantic defect labels, namely Missing_hole, Mouse_bite, Open_circuit, Short, Spur, and Spurious_copper, enabling the simulation of a realistic industrial environment. The data were partitioned into a "Seen" set (four classes: Mouse_bite, Open_circuit, Short, and Spur), which was used for training the baseline models, and a separate "Unseen" set (two classes: Missing_hole and Spurious_copper), which was reserved specifically for evaluating the system's few-shot and zero-shot adaptation capabilities.

Three primary approaches were evaluated: VLM-AOI (k=30), the proposed few-shot classifier that generates prototypes from 30 sample images for each of the six classes (180 samples in total); a baseline supervised ResNet-50 (k=30) trained using 30 samples from each of the four "Seen" classes (120 samples in total) to ensure a fair comparison; and a reference baseline, ResNet-50 (Full-data), trained using more

than 1,800 "Seen" images to serve as a maximum supervised-learning benchmark. Due to resource constraints, all experiments were conducted on the Google Colab platform using a GPU, with implementation performed in PyTorch and scikit-learn. The evaluated models included ResNet-50 (from torchvision) and VLM-AOI employing CLIP ViT-B/32.

### *B. Stage 1 (Anomaly Detection) Evaluation Protocol*

In practical industrial applications, traditional AOI systems are already highly effective at detecting anomalous regions. To ensure the evaluation focuses exclusively on the semantic classification power of the VLM and to maintain an objective comparison, a specific experimental design was adopted. Specifically, the ground-truth annotations (bounding boxes) from the DeepPCB dataset were used directly as the RoIs input for Stage 2. This approach effectively removed any influence of detection performance from the results, allowing for a pure and fair assessment of the classification capability, as all models were guaranteed to receive an identical set of RoIs.

### *C. Stage 2 (Defect Classification) Results*

This section presents the primary evaluation because it assesses the classification capability of the two models trained under identical few-shot settings (k=30) on both known and novel faults. It demonstrates the performance of both models when confronted with 940 unseen fault samples of Missing_hole and Spurious_copper, which were not included during training.

The results presented in Tables I and II offer clear evidence regarding the proposed models' ability to adapt to novel faults. The Baseline (ResNet-50 k=30) failed completely, achieving 0.00% accuracy. Its confusion matrix confirms that it randomly guessed all 940 "Unseen" images into the four "Seen" classes on which it was trained, for example, misclassifying 290 out of 467 Missing_hole images as Short. In sharp contrast, VLM-AOI demonstrated superior adaptation, achieving 82.9% overall accuracy. Notably, it identified the Missing_hole fault almost perfectly, with 98.3% recall, demonstrating its capability to learn distinct visual concepts effectively. Furthermore, VLM-AOI successfully overcame the domain gap for the technical concept of Spurious_copper; with just 30 sample images, it achieved 67.7% recall, demonstrating that the few-shot (image-based) approach is superior to zero-shot (text-based) learning.

The comparison of the two models on 1,833 "Seen" samples (known defects), as presented in Tables III and IV, where both models were trained using only 30 sample images per class, reveals notable differences in performance. Although the Baseline (ResNet-50 k=30) appears to significantly outperform VLM-AOI (61.3%), this outcome is expected. The high performance of the Baseline (ResNet-50 k=30) model, at 78.3%, stems from its function as a "rote-learning expert" in a narrow four-class classification task for which it is highly optimized. Conversely, the lower performance of VLM-AOI (k=30) at 61.3% is attributed to semantic ambiguity, as VLM-AOI must classify across six classes (including both Seen and Unseen categories). Specifically, the confusion matrix reveals that when the VLM encountered a Spur image, it incorrectly predicted Mouse_bite (120/458 times) and Spurious_copper (139/458 times) because all three concepts share semantic similarity ("an abnormal piece of copper"). The Baseline (ResNet-50) model does not face this issue because it is conceptually "blind" (i.e., it is unaware of the concept of Spurious_copper) and has simply rote learned the four classes.

TABLE I. COMPARISON OF PERFORMANCE ON THE UNSEEN SET

| Metric | Overall accuracy (%) | Precision | Recall | F1-score | Time (ms) |
|---|---|---|---|---|---|
| VLM-AOI (k=30) | 82.9 | 0.999 | 0.829 | 0.898 | 18.202 |
| Baseline (ResNet-50 k=30) | 0.00 | 0.000 | 0.000 | 0.000 | 14.535 |

TABLE II. DETAILED CLASS-WISE PERFORMANCE ON THE UNSEEN SET

| Class | Model | Precision | Recall | F1-score | Support (samples) |
|---|---|---|---|---|---|
| Missing_hole | VLM-AOI (k=30) | 0.998 | 0.983 | 0.990 | 467 |
| | Baseline (k=30) | 0.000 | 0.000 | 0.000 | 467 |
| Spurious_copper | VLM-AOI (k=30) | 1.000 | 0.677 | 0.807 | 473 |
| | Baseline (k=30) | 0.000 | 0.000 | 0.000 | 473 |

TABLE III. COMPARISON OF PERFORMANCE ON THE SEEN SET

| Metric | Overall accuracy (%) | Precision | Recall | F1-score | Time (ms) |
|---|---|---|---|---|---|
| VLM-AOI (k=30) | 61.3 | 0.704 | 0.613 | 0.643 | 16.036 |
| Baseline (ResNet-50 k=30) | 78.3 | 0.788 | 0.783 | 0.780 | 14.425 |

TABLE IV. DETAILED CLASS-WISE PERFORMANCE ON THE SEEN SET

| Class | Model | Precision | Recall | F1-score | Support (samples) |
|---|---|---|---|---|---|
| Mouse_bite | VLM-AOI (k=30) | 0.676 | 0.690 | 0.683 | 462 |
| | Baseline (k=30) | 0.807 | 0.652 | 0.721 | 462 |
| Open_circuit | VLM-AOI (k=30) | 0.696 | 0.652 | 0.669 | 452 |
| | Baseline (k=30) | 0.742 | 0.858 | 0.796 | 452 |
| Short | VLM-AOI (k=30) | 0.768 | 0.783 | 0.776 | 461 |
| | Baseline (k=30) | 0.779 | 0.918 | 0.843 | 461 |
| Spur | VLM-AOI (k=30) | 0.676 | 0.332 | 0.445 | 458 |
| | Baseline (k=30) | 0.822 | 0.707 | 0.761 | 458 |

The comparison highlights the main strength of the VLM-AOI framework: adaptability to novel defects (see Figure 3). The Baseline (ResNet-50), trained only on the Seen classes, completely fails on the Unseen classes, achieving 0% recall and revealing the brittleness of conventional supervised methods. In contrast, VLM-AOI achieves strong performance on these

unseen defects, including 98.3% recall for Missing_hole and 67.7% recall for Spurious_copper, demonstrating effective few-shot concept learning. However, this adaptability comes with a trade-off: VLM-AOI performs worse than the Baseline on Seen classes (e.g., 78.3% vs. 91.8% recall for Short). This gap reflects the challenge of semantic ambiguity, where broader concept learning can cause confusion between visually similar defect types.

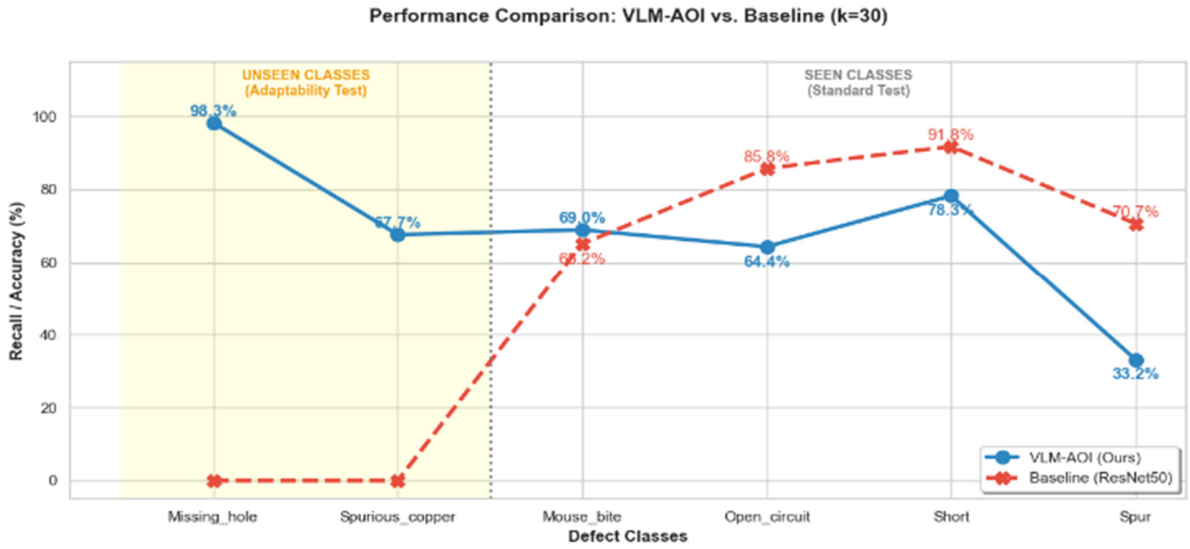


Fig. 3. Performance comparison between VLM-AOI and ResNet-50.

### D. *Qualitative Analysis: Visual Results of the VLM-AOI Framework*

This section provides a detailed qualitative analysis of the VLM-AOI performance through representative visual examples. The goal is to clarify how the model operates in real-world scenarios and to visually illustrate the challenges discussed in the quantitative results.

Figures 4–8 provide visual evidence of both the strengths and limitations of the proposed VLM-AOI framework. Figure 4 shows a successful adaptation to an unseen defect, where the model accurately localizes the RoI and classifies the Missing_hole class with high confidence, demonstrating strong concept-learning capability. Figures 5 and 7 illustrate successful recognition of seen defects such as Short and Spurious_copper, confirming the model's reliability when defect patterns are visually distinctive. In contrast, Figures 6 and 8 present failure cases caused by semantic ambiguity, where visually similar defects such as Mouse_bite and Spur are confused, and Short and Open_circuit are misclassified. This observation is consistent with the quantitative results in Tables I–IV and highlights a key limitation of the model.

Overall, the results confirm that VLM-AOI (few-shot) is a flexible and promising solution for adaptive defect classification. Few-shot learning effectively bridges the domain gap that causes zero-shot methods to fail on specialized industrial terms, as shown by the strong improvement on Spurious_copper. The model also introduces only a small increase in inference latency compared with the baseline, which is acceptable given its adaptability to new defects. However, semantic ambiguity remains the most important challenge, as the model may confuse visually similar defect concepts, leading to lower performance on the seen set compared to the baseline.

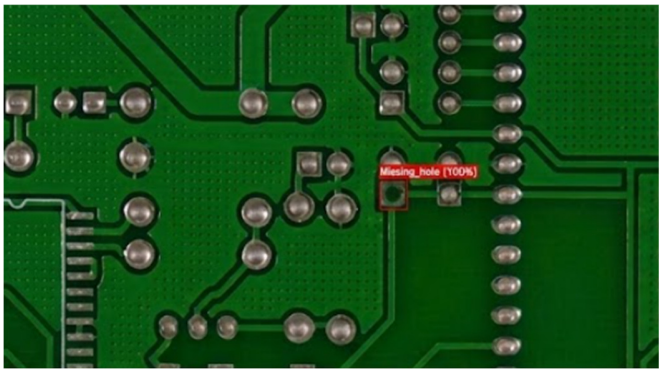


Fig. 4. Successful adaptation to Missing_hole defect.

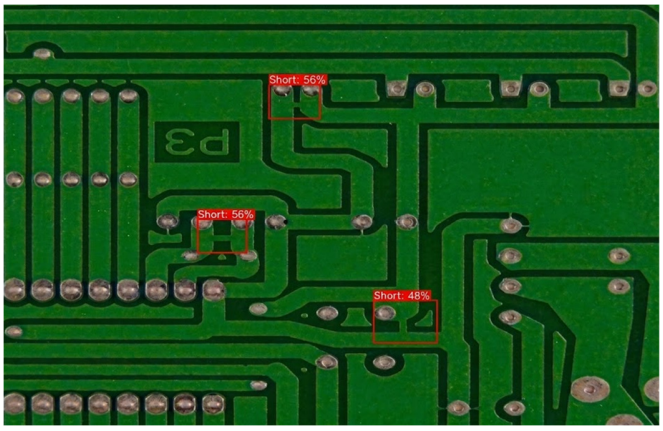


Fig. 5. Accurate classification of Short defect.

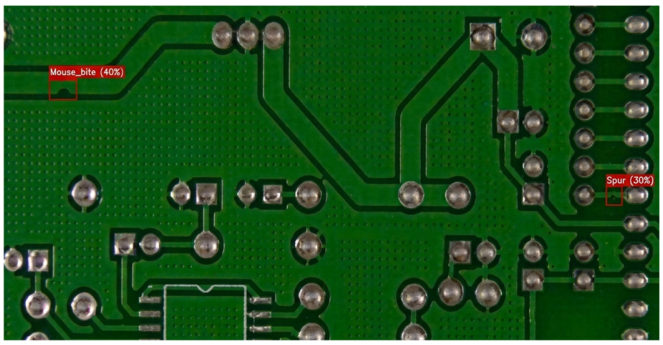


Fig. 6. Failure case: confusion between Mouse_bite and Spur.

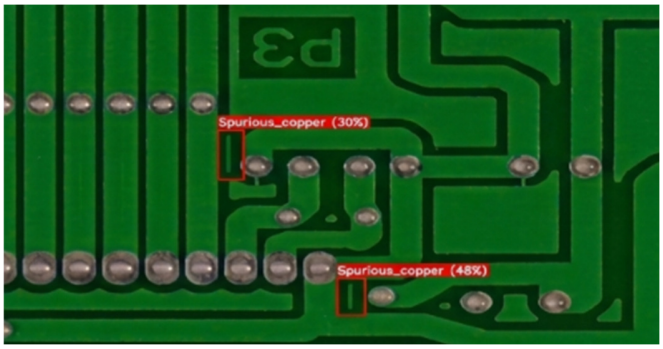


Fig. 7. Accurate classification of Spurious_copper defect.

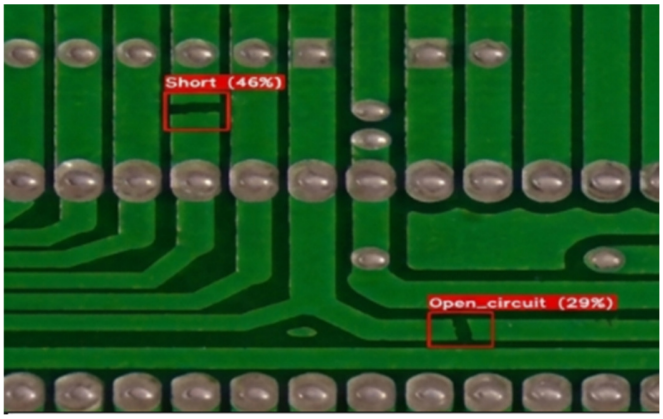


Fig. 8. Failure case: confusion between Short and Open_circuit.

## IV. CONCLUSION AND FUTURE WORK

This study introduces the VLM-AOI framework as a promising paradigm for next-generation autonomous quality assurance in smart manufacturing environments. By tightly integrating high-throughput unsupervised anomaly detection, represented by PatchCore, with the powerful semantic representation capability of large-scale Vision–Language Models (VLMs) such as Contrastive Language–Image Pretraining (CLIP), the proposed approach effectively mitigates the inherent "black-box" limitation of conventional Automated Optical Inspection (AOI) systems. In doing so, it bridges the gap between low-level visual inspection and high-level semantic reasoning.

Extensive quantitative evaluations reveal a substantial performance gap between traditional supervised approaches and the proposed method under out-of-distribution conditions. While the baseline classifier based on ResNet-50 exhibits complete failure (0.00% accuracy) when encountering unseen defect categories, the VLM-AOI framework achieves a notable accuracy of 82.9% on novel defect types such as Missing_hole and Spurious_copper. These results demonstrate strong cross-domain generalization capability, enabling the system to move beyond rigid pattern recognition toward a more adaptive and human-like interpretation of defect morphology. Importantly, this capability is achieved under limited annotation settings (k=30), highlighting the data efficiency of the proposed framework.

Despite these encouraging outcomes, the study also identifies semantic ambiguity as a key challenge. Defect classes with similar visual characteristics, such as Spur and Mouse_bite, can lead to degraded recall performance on in-distribution samples when compared with fully supervised baselines. This limitation underscores the need for more refined semantic alignment mechanisms between visual and textual representations.

Future research will therefore focus on enhancing multimodal feature alignment and developing domain-specific prompt engineering strategies tailored to industrial inspection tasks. In addition, efforts will be directed toward reducing computational complexity and optimizing inference efficiency, thereby facilitating real-time deployment on edge-AI platforms. Such advancements are essential for ensuring the scalability and practical adoption of VLM-based AOI systems in increasingly dynamic and data-driven manufacturing ecosystems.

## DECLARATION OF COMPETING INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## ACKNOWLEDGMENT

The authors would like to thank the Electric Power University for the institutional support and infrastructure provided during this research.

## DATA AVAILABILITY

The DeepPCB dataset utilized in this research is a publicly available benchmark and can be accessed at [18].

## REFERENCES

[1] S. Ruengrote, K. Kasetravetin, P. Srisom, T. Sukchok, and D. Kaewdook, "Design of Deep Learning Techniques for PCBs Defect Detecting System based on YOLOv10," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 18741–18749, Dec. 2024, https://doi.org/10.48084/etasr.9028.

[2] N. Sajitha and S. P. Priya, "Optimal Artificial Neural Network-based Fabric Defect Detection and Classification," *Engineering, Technology & Applied Science Research*, vol. 14, no. 2, pp. 13148–13152, Apr. 2024, https://doi.org/10.48084/etasr.6773.

[3] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. Gehler, "Towards Total Recall in Industrial Anomaly Detection," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*,

New Orleans, LA, USA, 2022, pp. 14298–14308, https://doi.org/10.1109/CVPR52688.2022.01392.

[4] T. Defard, A. Setkov, A. Loesch, and R. Audigier, "PaDiM: A Patch Distribution Modeling Framework for Anomaly Detection and Localization," in *25th International Conference on Pattern Recognition Workshops*, Virtual Event, 2021, pp. 475–489, https://doi.org/10.1007/978-3-030-68799-1_35.

[5] P. Bergmann, K. Batzner, M. Fauser, D. Sattlegger, and C. Steger, "The MVTec Anomaly Detection Dataset: A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection," *International Journal of Computer Vision*, vol. 129, no. 4, pp. 1038–1059, Apr. 2021, https://doi.org/10.1007/s11263-020-01400-4.

[6] M. A. M. Sathiaseelan, O. P. Paradis, S. Taheri, and N. Asadizanjani, "Why Is Deep Learning Challenging for Printed Circuit Board (PCB) Component Recognition and How Can We Address It?," *Cryptography*, vol. 5, no. 1, Mar. 2021, Art. no. 9, https://doi.org/10.3390/cryptography5010009.

[7] Y. Liu, C. Zhang, and X. Dong, "A survey of real-time surface defect inspection methods based on deep learning," *Artificial Intelligence Review*, vol. 56, no. 10, pp. 12131–12170, Oct. 2023, https://doi.org/10.1007/s10462-023-10475-7.

[8] W. Wang, V. W. Zheng, H. Yu, and C. Miao, "A Survey of Zero-Shot Learning: Settings, Methods, and Applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, Jan. 2019, Art. no. 13, https://doi.org/10.1145/3293318.

[9] A. Radford *et al.*, "Learning Transferable Visual Models From Natural Language Supervision," in *Proceedings of the 38th International Conference on Machine Learning*, Virtual Event, 2021, pp. 8748–8763.

[10] C. Jia *et al.*, "Scaling Up Visual and Vision-Language Representation Learning With Noisy Text Supervision," in *Proceedings of the 38th International Conference on Machine Learning*, Virtual Event, 2021, pp. 4904–4916.

[11] K. Moenck, D. T. Thieu, J. Koch, and T. Schüppstuhl, "Industrial Language-Image Dataset (ILID): Adapting Vision Foundation Models for Industrial Settings," *Procedia CIRP*, vol. 130, pp. 250–263, Jan. 2024, https://doi.org/10.1016/j.procir.2024.10.084.

[12] J. Zhou, W. Liu, G. Yang, H. Zhao, and F. Yuan, "Prompting Industrial Anomaly Segment with Large Vision-Language Models," in *Proceedings of the 6th ACM International Conference on Multimedia in Asia*, Auckland, New Zealand, 2024, https://doi.org/10.1145/3696409.3700192.

[13] G. Shinde, A. Ravi, E. Dey, S. Sakib, M. Rampure, and N. Roy, "A Survey on Efficient Vision-Language Models," *WIREs Data Mining and Knowledge Discovery*, vol. 15, no. 3, Sept. 2025, Art. no. e70036, https://doi.org/10.1002/widm.70036.

[14] Z. Li, Y. Yan, X. Wang, Y. Ge, and L. Meng, "A survey of deep learning for industrial visual anomaly detection," *Artificial Intelligence Review*, vol. 58, no. 9, June 2025, Art. no. 279, https://doi.org/10.1007/s10462-025-11287-7.

[15] Q. Ling and N. A. M. Isa, "Printed Circuit Board Defect Detection Methods Based on Image Processing, Machine Learning and Deep Learning: A Survey," *IEEE Access*, vol. 11, pp. 15921–15944, 2023, https://doi.org/10.1109/ACCESS.2023.3245093.

[16] J. Jeong, Y. Zou, T. Kim, D. Zhang, A. Ravichandran, and O. Dabeer, "WinCLIP: Zero-/Few-Shot Anomaly Classification and Segmentation," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, Canada, 2023, pp. 19606–19616, https://doi.org/10.1109/CVPR52729.2023.01878.

[17] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," in *Proceedings of the 37th International Conference on Machine Learning*, Virtual Event, 2020, pp. 1597–1607.

[18] tangsanli5201, "tangsanli5201/DeepPCB." May 13, 2026. [Online]. Available: https://github.com/tangsanli5201/DeepPCB.

[19] S. Tang, F. He, X. Huang, and J. Yang, "Online PCB Defect Detector On A New PCB Defect Dataset." arXiv, Feb. 17, 2019, https://doi.org/10.48550/arXiv.1902.06197.